From Cognitive Extension to Cognitive Dependence

A History Theory and Governance of Human Offloading in the Age of Generative AI

Frank D. Smith, B.S., MCSE

September 4, 2026

Version 1.1

Abstract

Humans have always altered their environments to reduce the burden of remembering, calculating, navigating, and deciding. Contemporary psychology calls this behavior *cognitive offloading*: the use of external action or resources to reduce the internal information-processing demands of a task. Clay tokens, writing, diagrams, books, calculators, spellcheckers, search engines, satellite navigation, voice assistants, and generative artificial intelligence (AI) belong to a long—but discontinuous—history of such practices. Offloading is not intrinsically harmful.

It expands effective memory, permits coordination at scales unavailable to unaided minds, improves immediate performance, and can release attention for higher-order work. The danger arises when assistance routinely substitutes for the cognitive operations through which knowledge, skill, judgment, and error-detection capacity are formed. Generative AI intensifies this danger because it can offload not merely storage or calculation but the apparent performance of synthesis, explanation, composition, and reasoning. This paper traces the intellectual and technological history of cognitive offloading, identifies the mechanisms that encourage it, distinguishes productive cognitive extension from dependency and deskilling, and proposes remedies in education, interface design, professional practice, and public policy.

Drawing on game theory, it rejects the assumption that increased machine capability must produce a corresponding loss of human capability. Cognitive sovereignty can change the structure of human–AI interaction from substitution toward complementarity, making positive-sum outcomes possible, although not automatic. The paper further argues that cognitive atrophy should not be presented as a numerically proven alternative to AI-caused catastrophe. Rather, it is an observable, nearer-term risk that may compound other AI risks by degrading the human expertise and institutional capacity required to detect error, exercise oversight, and make meaningful use of increasingly powerful tools.

Keywords: cognitive offloading; generative artificial intelligence; extended mind; distributed cognition; automation bias; metacognition; deskilling; critical thinking

1 Introduction

The most consequential effect of artificial intelligence may not be that machines acquire human capacities, but that humans cease regularly to exercise them. A person who once recalled a telephone number consults a contact list; one who once estimated a sum opens a calculator; one who once navigated from landmarks follows turn-by-turn directions; and one who once formulated a question, searched sources, compared claims, and composed an answer may now ask a conversational system for a finished response. These cases differ in both degree and kind, yet all redistribute work between a person and an external resource.

Psychologists Evan Risko and Sam Gilbert gave the contemporary field its canonical synthesis. They define cognitive offloading as the use of physical action to alter a task's information-processing requirements and thereby reduce cognitive demand (Risko & Gilbert, 2016). Their article did not invent the underlying phenomenon, and the exact first use of the phrase remains unclear. It consolidated research previously conducted under such labels as external memory, cognitive artifacts, distributed cognition, transactive memory, and computational offloading. Accordingly, Risko and Gilbert should be credited with establishing the standard modern account, not uncritically described as the phrase's sole originators.

This paper advances three claims. First, offloading is a constitutive feature of human cognition and civilization, not a pathology unique to digital technology. Second, the benefits of offloading coexist with a substitution effect: when a tool replaces the effort that produces learning or maintains expertise, immediate performance can improve while later unaided capacity declines. This vulnerability is aggravated by an educational gap: critical thinking is frequently named as a curricular aspiration but is not consistently made an explicit, practiced, and assessed object of instruction. Third, generative AI marks a qualitative escalation because it offers to perform open-ended cognitive products—arguments, interpretations, plans, judgments, and prose—while concealing much of the process by which those products were generated. The appropriate response is neither a ban on cognitive tools nor passive acceptance, but the deliberate design of systems and norms that preserve human agency, learning, and fallback competence.

2 What Cognitive Offloading Is and Is Not

Cognitive offloading includes simple bodily strategies, such as rotating a page rather than mentally rotating an image; environmental modifications, such as placing an object by a door as a reminder; symbolic artifacts, such as lists and diagrams; and digital systems that store, retrieve, calculate, recommend, or generate information. People tend to offload more when tasks are difficult, when they doubt their memory, and when external action is cheap and reliable. Even young children invent external strategies and use them more frequently as task difficulty rises (Risko & Gilbert, 2016).

Offloading must be distinguished from three related ideas. *Cognitive load* describes demands placed on limited working-memory resources; offloading is one strategy for changing those demands. *Automation* transfers the execution of a function to a machine, but need not reduce meaningful human thought—for example, automating transcription may allow closer analysis. *Cognitive extension* is the stronger philosophical claim that a reliably integrated external resource can become part of a person's cognitive system. Clark and Chalmers (1998) illustrate this with a notebook that functions for its user much as biological memory does. These concepts overlap, but none implies that tool use is necessarily harmful.

The normative question is therefore not whether cognition is internal or external. It is whether the combined human-tool system preserves the capacities needed for its own direction, verification, and recovery. Offloading is *augmentative* when it removes low-value burden while the user retains understanding and control. It becomes *substitutive* when it bypasses the operations through which the user would acquire or sustain knowledge. It becomes *dependent* when the user cannot perform, evaluate, or responsibly supervise the task without the tool.

3 A Brief History of Externalized Cognition

3.1 Gesture objects counting tokens and writing

The history of offloading predates electronic devices and probably predates writing. Fingers, marks, knotted cords, arranged objects, and other material representations allow quantities and commitments to remain stable outside working memory. Archaeologist Denise Schmandt-Besserat argues that Near Eastern clay tokens, appearing by approximately 7500 BCE, represented units of goods and contributed over millennia to the development of numerical abstraction and writing. Markings on clay envelopes later represented the tokens contained within them, and early tablets emerged in the late fourth millennium BCE (Schmandt-Besserat, 1979, 2013). Although scholars dispute elements of a single linear token-to-writing account, the artifacts clearly illustrate how external representations can both store information and enable new abstractions.

Writing is the foundational case because it demonstrates the dual nature of offloading. In Plato's *Phaedrus*, Socrates recounts the myth of Theuth, whose invention of writing is judged to provide reminders rather than genuine memory and to give learners the appearance rather than the reality of wisdom (Plato, ca. 360 BCE/1925, 274c–275b). The passage is not a modern experiment and should not be treated as one. It is, however, an extraordinarily early articulation of the substitution concern: an external aid may expand access to information while reducing internal practice and encouraging unearned confidence. History also exposes the limit of the complaint. Writing did not simply weaken memory; it enabled law, cumulative scholarship, science, administration, and criticism across time. The same artifact could displace one ability while making previously impossible cognition available.

3.2 Manuscripts print tables maps and mechanical calculation

Indexes, marginalia, commonplace books, diagrams, multiplication tables, maps, and printed reference works did more than preserve facts. They reorganized problems so that perception could replace recall and spatial arrangement could replace mental manipulation. Such artifacts are not passive containers. In distributed-cognition accounts, people, representations, and procedures form a functional system in which information is transformed across media and participants (Hutchins, 1995). A ship's navigation, for example, is accomplished not by a solitary mind but by coordinated people, instruments, charts, and conventions.

The abacus, slide rule, mechanical calculator, and electronic calculator progressively externalized arithmetic operations. Yet the lesson is not a simple story of decline. An abacus can serve as both an external calculating device and a training system whose expert users internalize a visuospatial representation. Tools can therefore scaffold capacity before replacing performance. Whether a device educates or deskills depends on when it is introduced, what the user must understand, and whether unaided practice continues.

3.3 Spellcheck search photography and satellite navigation

Personal computing lowered the cost of offloading and embedded it in ordinary action. Spellcheckers externalized error detection; calendars and notifications externalized prospective memory; digital photography externalized episodic recording; and search engines externalized factual retrieval. In four experiments, Sparrow, Liu, and Wegner (2011) found that expecting future computer access reduced memory for information while improving memory for where it could be found. They interpreted the Internet as a form of external or transactive memory. This was a redistribution rather than the disappearance of memory, but it left users reliant on continued access and able to know where without necessarily knowing what.

The same pattern appears elsewhere. Museum visitors who photographed whole objects later remembered fewer objects and details than visitors who only observed them, although focused photography reduced this impairment (Henkel, 2014). Habitual GPS use has been associated with poorer spatial-memory performance during self-guided navigation, with greater GPS use over time associated with steeper decline in one longitudinal study (Dahmani & Bohbot, 2020). Such findings do not establish universal, irreversible brain deterioration. They show that attention, expectation, and repeated strategy choice shape what is encoded and practiced.

3.4 Voice assistants and generative AI

Search engines still generally require a user to formulate queries, inspect multiple results, judge sources, and assemble an answer. Voice assistants reduce even the small friction of opening a browser and scanning a page. Their significance is less the complexity of any single request than the formation of a default: uncertainty is followed immediately by external query rather than recall, estimation, or reflection.

Generative AI goes further. It can produce a coherent-looking endpoint for tasks whose educational or professional value lies partly in the process: outlining a problem, selecting evidence, reconciling contradictions, composing an explanation, and revising a judgment. The output's fluency can make the transferred cognition difficult to see. A calculator visibly returns a number; a language model produces an explanation that simulates the linguistic signs of understanding while sometimes fabricating facts or reasoning incorrectly.

Evidence about long-term cognitive effects remains young. A 2025 study of knowledge workers found that greater confidence in generative AI was associated with less self-reported critical-thinking effort, while greater confidence in one's own ability was associated with more critical engagement (Lee et al., 2025). This is not proof that AI use causes generalized loss of intelligence: much of the current evidence is correlational, self-reported, task-specific, or based on small samples. The careful conclusion is that generative AI creates unusually broad opportunities for substitution, and early evidence is consistent with concern about reduced engagement. Strong longitudinal and preregistered experiments are still needed.

4 Why Humans Offload

4.1 Bounded working memory and rational effort allocation

Working memory and attention are limited. Externalizing intermediate steps can improve accuracy and free resources for planning. From the individual's perspective, offloading is often rational: if retrieval is uncertain and a trusted answer is seconds away, consulting a device minimizes time and error. Risko and Gilbert (2016) describe offloading choices as sensitive to the relative costs and benefits of internal and external strategies.

4.2 Metacognitive confidence

People decide whether to think internally partly by estimating their own capacity. Low confidence can prompt useful support-seeking, but inaccurate confidence can produce two opposite failures: unnecessary dependence or unjustified refusal of assistance. Tool fluency complicates this calibration because a high-quality output can be misread as evidence of the user's own mastery.

4.3 The economics of convenience

Digital services compete to eliminate friction. The fewer steps between a desire and a result, the more likely the external strategy is to win the user's implicit cost comparison. Subscription models, engagement incentives, workplace productivity metrics, and educational pressure all reward the visible product more readily than the invisible cognition that produced it. Offloading is therefore not merely a personal failure of discipline; it is an economically encouraged behavior.

4.4 Trust automation bias and opacity

Parasuraman and Riley (1997) describe automation *misuse* as overreliance that can yield failures of monitoring and biased decisions. With generative systems, polished language, rapid response, and uncertain provenance can magnify this tendency. A person with little domain knowledge is least equipped to recognize a plausible error, yet may have the strongest incentive to delegate. This creates an epistemic paradox: the tool is most useful where the user is weak, but responsible supervision requires some of the very knowledge being bypassed.

4.5 Skill acquisition requires effort

The ease of present performance is a poor proxy for durable learning. Research on “desirable difficulties” shows that effortful retrieval, generation, spacing, and interleaving can slow or impair immediate performance while improving later retention and transfer (Bjork & Bjork, 2011). If AI supplies an answer before a learner attempts retrieval or generation, it can remove the event that would have strengthened the learner. The relevant mechanism is use-dependent adaptation, not a moralized notion of laziness.

4.6 The critical thinking instruction gap

Critical thinking is often invoked as a general purpose of education but less often translated into a visible sequence of concepts, modeled reasoning, guided practice, feedback, and assessment. This distinction matters. A curriculum may ask students to “analyze” or “evaluate” material without teaching them how to distinguish a premise from a conclusion, test the relevance and credibility of evidence, identify assumptions, compare alternative explanations, calibrate confidence, or recognize common fallacies. Students are then expected to acquire critical thinking incidentally through subject-matter study.

It would be too broad to claim that schools generally ignore critical thinking. An OECD curriculum-mapping exercise found critical thinking represented frequently in the intended curricula of participating countries and jurisdictions. Yet the OECD also reports that what it means to develop and assess these skills in classroom practice often remains unclear (Vincent-Lancrin et al., 2019, 2020). The defensible diagnosis is an *implementation gap*: rhetorical inclusion is widespread, while explicit, systematic instruction and direct assessment are uneven.

This gap is consequential because critical thinking can be taught. Abrami et al.'s (2015) meta-analysis synthesized 341 effect sizes from experimental and quasi-experimental studies and found a positive average effect of critical-thinking instruction. Dialogue, authentic or situated problems, and mentoring were among the effective features. Earlier and subsequent reviews likewise favor approaches in which thinking concepts and skills are explicitly identified and practiced rather than merely expected to emerge (Bangert-Drowns & Bankert, 1990; Nsangi et al., 2025). The effects are heterogeneous and transfer across domains cannot be assumed, but the evidence rejects the idea that critical thinking is an innate trait or an automatic by-product of education.

The arrival of generative AI makes the implementation gap more dangerous. A student who has not been explicitly taught how claims are structured or evidence is evaluated is poorly positioned to interrogate a fluent synthetic answer. AI does not create this educational weakness; it scales the number of occasions on which the weakness matters. Critical-thinking education is therefore a direct counterforce to harmful cognitive offloading: it supplies both the skills needed to verify external cognition and the disposition to pause before accepting it.

5 Benefits Harms and the Threshold of Dependence

Offloading produces genuine benefits. It reduces error in routine calculation, provides accessibility for people with cognitive or sensory disabilities, coordinates large organizations, and permits individuals to work beyond biological memory limits. Studies of offloading often find improved immediate task performance even when later memory for the offloaded material is worse (Grinschgl et al., 2021). Any account that treats unaided cognition as inherently superior ignores the fact that science, engineering, and government depend on external representations.

The harms arise in at least five forms:

1. **Encoding loss:** information expected to remain available may receive shallower processing and be remembered less well.
2. **Skill decay:** procedures not practiced become slower, less accurate, or unavailable, especially under failure conditions.
3. **Verification loss:** reduced domain knowledge makes machine errors harder to detect.
4. **Agency loss:** users may allow defaults, recommendations, or generated framings to determine goals rather than merely help achieve them.
5. **Systemic fragility:** institutions that optimize away apprenticeships and manual fallback capacity may perform efficiently in normal conditions but fail catastrophically when tools are unavailable or wrong.

The threshold between augmentation and dependence can be tested with three questions: Can the user explain the result? Can the user detect a materially wrong result? Can the user continue at an acceptable minimum level if the tool fails? The answers need not always be yes. Few people should be required to reproduce a modern processor's arithmetic by hand. But where errors affect rights, safety, knowledge, or governance, retaining human comprehension and fallback competence is essential.

6 From Zero Sum Substitution to Positive Sum Complementarity

6.1 The perceived zero sum game

Much public discussion implicitly treats human and machine cognition as a fixed pool: every capability transferred to AI is a capability lost by people, while every restriction placed on AI preserves a corresponding amount of human agency. This is a zero-sum mental model, even when the term is not used. It appears both in fears that AI must defeat or replace humanity and in the assumption that any use of AI necessarily weakens its user.

Game theory cautions against treating all interactions as contests with a fixed total payoff. In a zero-sum game, one player's gain is matched by another's loss. Many social and economic interactions instead permit mutual gains, mutual losses, or a mixture of cooperation and competition. Fisher (2008) emphasizes this broader side of game theory: its usefulness for understanding how rules and incentives can trap people in collectively inferior outcomes or enable cooperation. The relevant question is not simply who “wins,” human or machine, but how the rules of interaction shape the attainable payoffs.

Strictly speaking, ordinary AI assistance is not always a game between a human and an AI. Game-theoretic analysis requires actors with strategies and payoffs; a non-agentic tool has neither independent welfare nor necessarily an independent objective. The strategic actors may instead be users, employers, schools, developers, platforms, and governments, each responding to the choices of the others. More autonomous AI systems can be modeled as agents to the extent that they pursue objectives and adapt strategically. This distinction prevents the game metaphor from substituting for analysis.

6.2 The cognitive commons

Cognitive capacity can be understood as both a personal resource and a social commons. An individual receives an immediate private benefit by delegating a difficult task: saved time, reduced effort, or a polished product. The cost of diminished practice may be delayed and partly borne by others—future employers, patients, students, voters, institutions, and the next generation of trainees. If everyone is rewarded for the immediate product while no one is rewarded for maintaining the underlying capacity, individually rational offloading can produce a collectively inferior equilibrium.

This resembles the cooperation problems Fisher (2008) describes through the tragedy of the commons and prisoner's dilemma. A student may reason that one AI-completed assignment will not damage the educational system; a firm may reason that eliminating one cohort of junior tasks will not exhaust the future supply of experts; and a platform may reason that one more frictionless answer improves user welfare. When these strategies become general, however, the shared stock of practiced judgment and expertise may decline. The “cognitive commons” is not literally finite in the manner of a pasture: it can grow through education and practice. That makes its depletion avoidable, but only if institutions recognize the externality and reward its renewal.

6.3 Cognitive sovereignty as a positive sum strategy

Cognitive sovereignty rejects both isolation and surrender. It is the maintained capacity to set goals, form an initial judgment, understand relevant evidence, detect error, and remain capable when an external system fails. A sovereign user can delegate selectively while preserving responsibility. Under those conditions, AI may contribute speed, scale, retrieval, simulation, pattern detection, and alternative generation, while the human contributes situated goals, causal understanding, moral and political judgment, accountability, and the ability to question the frame of the task. The total capability of the combined system may then exceed either component working alone.

This is a positive-sum hypothesis, not an established universal result. Vaccaro, Almaatouq, and Malone's (2024) meta-analysis of 106 experiments found that human–AI combinations generally improved on unaided human performance but, on average, failed to outperform the better of the human or AI acting alone. Complementarity varied by task: combinations showed gains more often in content-creation tasks and losses more often in decision tasks. Notably, combinations tended to gain when humans alone were stronger than AI and to lose when AI alone was stronger. These results show that placing a human "in the loop" does not itself create synergy.

Cognitive sovereignty supplies a proposed explanation for when positive-sum performance may be achievable. Complementarity requires a human who possesses an independent signal, understands the task, recognizes the system's comparative weaknesses, and has both authority and incentive to override it. A user who merely approves a generated answer contributes little additional information and may introduce automation bias. Conversely, an AI system used only as a passive lookup table forfeits its generative advantages. Positive-sum design assigns work according to complementary strengths while preserving enough overlap for mutual error correction.

6.4 Changing the payoff structure

The present equilibrium often rewards completion speed and output volume while treating human learning, verification, and fallback competence as uncompensated costs. Cognitive-sovereignty interventions change those payoffs. Schools can grade reasoning traces and independent verification rather than polished answers alone. Employers can measure error detection and preserve apprenticeship instead of rewarding only throughput. AI interfaces can make users generate an initial answer, expose uncertainty, and reward successful challenges. Regulators and professional bodies can impose accountability that cannot be delegated to a model.

These changes do not require the human to perform every operation. They make retained judgment valuable within the system. In game-theoretic terms, the objective is to alter the rules so that maintaining human capability is individually rational as well as collectively beneficial. The desired equilibrium is not humans competing against AI for a fixed share of cognition, but humans and machines producing a larger surplus without consuming the human capacities required to direct and evaluate it.

7 Cognitive Offloading and Existential AI Risk

Public discussion often concentrates on a future "loss of control" in which highly capable systems operate beyond meaningful human command. That risk should not be dismissed. In a 2023 survey of 2,778 researchers publishing in leading AI venues, respondents assigned widely varying probabilities to extremely bad outcomes, illustrating substantial uncertainty rather than a settled forecast (Grace et al., 2024). The *International AI Safety Report 2026* likewise concludes that current systems lack the capabilities required for loss-of-control scenarios while noting progress in relevant capabilities and serious uncertainty (Bengio et al., 2026).

It would therefore be academically indefensible to assert that cognitive offloading has been proven more probable than AI-caused extinction: the outcomes, time horizons, and metrics are not commensurable, and evidence for long-term generalized cognitive decline is incomplete. A stronger claim is available. Cognitive substitution is already occurring, is open to empirical measurement, and may increase gradually without a salient disaster. It can also act as a *risk multiplier*. Technical and democratic control of AI depends upon people who can formulate goals, understand systems, contest outputs, maintain independent knowledge, and recognize manipulation. If those capacities erode, society loses not only particular skills but part of the oversight layer intended to prevent other AI harms.

The question "If AI took over, how would we know?" should thus be translated from rhetoric into operational criteria. Control has weakened when humans no longer set consequential goals; cannot obtain intelligible accounts of system behavior; cannot audit inputs, processes, and outcomes; routinely accept recommendations without independent checks; or lack the expertise and infrastructure to suspend or replace the system. A dramatic machine rebellion is unnecessary. Functional loss of control can emerge incrementally when institutions retain nominal authority but lose practical epistemic independence.

8 Remedies Preserving the Human in the Loop and in the Mind

8.1 Teach critical thinking explicitly repeatedly and within disciplines

Schools should make the operations of critical thought visible: identifying claims and reasons; distinguishing deduction, induction, and abduction; evaluating source credibility; detecting ambiguity and hidden assumptions; testing causal claims; generating alternative explanations; recognizing fallacies and cognitive biases; and calibrating confidence. Instruction should combine a shared vocabulary with repeated application inside history, science, mathematics, media, and civic problems. Generic lessons alone risk weak transfer; immersion alone risks leaving the underlying operations unnamed.

Assessment must match the goal. Students should be asked to annotate arguments, compare conflicting sources, explain why evidence changes a conclusion, expose the weakest step in a generated answer, and revise beliefs in response to counterevidence. These performances should be assessed both with and without AI assistance. Unaided assessment measures retained capacity; assisted assessment measures whether the learner can use a tool without surrendering judgment.

Accessible public resources can complement formal curricula. The nonprofit School of Thought, for example, publishes Creative Commons materials that translate logical fallacies, cognitive biases, and creative-thinking strategies into concise visual explanations (School of Thought, n.d.). Its widely circulated fallacy materials illustrate how reasoning vocabulary can be made memorable and usable. Such materials are best treated as instructional resources, not as empirical evidence of effectiveness; they should be paired with guided practice, authentic examples, and research-supported pedagogy.

8.2 Use a cognition first sequence

For learning and judgment tasks, users should attempt recall, estimation, diagnosis, or an initial outline before consulting AI. The tool should then critique, question, or compare rather than immediately replace the attempt. This preserves generation and retrieval practice while still providing feedback.

8.3 Design AI as a tutor and adversarial collaborator

Interfaces can default to questions, hints, partial steps, counterarguments, and requests for justification. A “challenge mode” could withhold a complete answer until the user records an initial view. A “verification mode” could insert plausible errors for supervised detection or require users to identify the evidence supporting each claim. The aim is not arbitrary friction but calibrated, desirable difficulty.

8.4 Establish cognitive fitness floors

Schools and safety-critical professions should identify competencies that must remain available without automation. Periodic unaided assessments can cover mental estimation, source evaluation, spatial orientation, diagnostic reasoning, writing, and emergency procedures as appropriate to the role. These floors should be justified by actual fallback and oversight needs, not nostalgia.

8.5 Separate practice from performance

During instruction, some tasks should prohibit or constrain AI because the task's purpose is to build the capacity being automated. During authentic performance, assistance may be appropriate because accuracy and efficiency are the goals. Policies should state which mode applies. Treating every assignment as a finished product invites substitution; treating every tool as cheating forfeits useful augmentation.

8.6 Require process visibility and provenance

For consequential uses, users should disclose what was delegated, preserve sources, and document independent checks. Systems should distinguish retrieved evidence from generated synthesis, expose uncertainty, and support comparison with primary sources. Process records cannot prove understanding, but they make invisible delegation more auditable.

8.7 Preserve apprenticeship and manual fallback

Organizations should not automate all junior work merely because experts can supervise AI-generated products today. Entry-level tasks often create the experience from which future experts acquire pattern recognition and judgment. Workforce planning should treat expertise as renewable infrastructure: it requires a pipeline, practice opportunities, and periodic exercise.

8.8 Measure outcomes longitudinally

Research and policy should track unaided retention, transfer, error detection, confidence calibration, and recovery after tool failure—not only AI-assisted task speed. Priority studies include longitudinal cohorts, randomized comparisons of answer-first and question-first interfaces, dose-response studies of habitual use, and research across ages, cultures, abilities, and occupational domains.

8.9 Adopt a personal delegation rule

At the individual level, a useful rule is: **offload storage and repetition readily; offload judgment cautiously; never outsource goals or final responsibility.** Before accepting an AI output, users can ask: What do I already think? What evidence would change my mind? What could this system be missing? Could I explain and defend the result without repeating its wording?

9 Conclusion

Human intelligence has never been confined to the skull. Civilization advances by constructing external memories, representations, and machines that allow finite minds to coordinate and discover more than any individual could alone. The relevant choice is therefore not between "natural" thought and technological corruption. It is between forms of human-tool organization that cultivate judgment and forms that quietly make judgment unnecessary until it is unavailable.

Generative AI sharpens an ancient dilemma because it can supply the outward products of cognition at unprecedented breadth, speed, and plausibility. Present evidence does not justify claims that AI has already caused a population-wide collapse in intelligence, nor does it allow a numerical comparison between cognitive decline and extinction risk. It does justify precaution: offloading can improve immediate performance while weakening memory, engagement, or skill; automation can induce overreliance; and weak human expertise makes faulty automation harder to supervise.

The best remedy is not abstinence but cognitive sovereignty: the maintained ability to set ends, form independent representations, interrogate evidence, detect error, and continue when the tool fails. This is also the condition that can turn AI from a perceived zero-sum competitor into a positive-sum complement. The surplus is real only when the combined system exceeds what either participant could accomplish alone without liquidating the human expertise needed to direct it. Super-tools are valuable only to people and institutions capable of deciding what should be done with them. Preserving that capability is not separate from AI safety. It is one of its preconditions.

References

Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research, 85*(2), 275–314. https://doi.org/10.3102/0034654314551063

Bangert-Drowns, R. L., & Bankert, E. (1990, April). *Meta-analysis of effects of explicit instruction for critical thinking* [Conference paper]. American Educational Research Association Annual Meeting, Boston, MA. https://eric.ed.gov/?id=ED328614

Bengio, Y., et al. (2026). *International AI safety report 2026*. International AI Safety Report. https://internationalaisafetyreport.org/publication/international-ai-safety-report-2026

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth.

Clark, A., & Chalmers, D. J. (1998). The extended mind. *Analysis, 58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Dahmani, L., & Bohbot, V. D. (2020). Habitual use of GPS negatively impacts spatial memory during self-guided navigation. *Scientific Reports, 10*, Article 6310. https://doi.org/10.1038/s41598-020-62877-0

Fisher, L. (2008). *Rock, paper, scissors: Game theory in everyday life*. Basic Books.

Grace, K., Stewart, H., Sandkühler, J. F., Thomas, S., Weinstein-Raun, B., & Brauner, J. (2024). Thousands of AI authors on the future of AI. *arXiv*. https://doi.org/10.48550/arXiv.2401.02843

Grinschgl, S., Meyerhoff, H. S., & Papenmeier, F. (2021). Consequences of cognitive offloading: Boosting performance but diminishing memory. *Quarterly Journal of Experimental Psychology, 74*(9), 1477–1496. https://doi.org/10.1177/17470218211008060

Henkel, L. A. (2014). Point-and-shoot memories: The influence of taking photos on memory for a museum tour. *Psychological Science, 25*(2), 396–402. https://doi.org/10.1177/0956797613504438

Hutchins, E. (1995). *Cognition in the wild*. MIT Press.

Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. https://doi.org/10.1145/3706598.3713778

Nsangi, A., et al. (2025). The effects of teaching strategies on learning to think critically in primary and secondary schools: An overview of systematic reviews. *Systematic Reviews, 14*. https://pmc.ncbi.nlm.nih.gov/articles/PMC12022541/

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Plato. (1925). *Phaedrus* (H. N. Fowler, Trans.). Harvard University Press. (Original work composed ca. 360 BCE). https://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A1999.01.0174%3Atext%3DPhaedrus%3Apage%3D275

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Schmandt-Besserat, D. (1979). An archaic recording system in the Uruk-Jemdet Nasr period. *American Journal of Archaeology, 83*(1), 19–48. https://doi.org/10.2307/504234

Schmandt-Besserat, D. (2013). Tokens and writing: The cognitive development. In D. Schmandt-Besserat (Ed.), *Symbols at 'Ain Ghazal* (pp. 56–62). Ex Oriente. https://sites.utexas.edu/dsb/tokens/tokens-and-writing-the-cognitive-development/

School of Thought. (n.d.). *School of Thought*. Retrieved September 4, 2026, from https://www.schoolofthought.org/

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science, 333*(6043), 776–778. https://doi.org/10.1126/science.1207745

Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour, 8*, 2293–2303. https://doi.org/10.1038/s41562-024-02024-1

Vincent-Lancrin, S., González-Sancho, C., Bouckaert, M., de Luca, F., Fernández-Barrerra, M., Jacotin, G., Urgel, J., & Vidal, Q. (2019). *Fostering students' creativity and critical thinking: What it means in school*. OECD Publishing. https://doi.org/10.1787/62212c37-en

Vincent-Lancrin, S., et al. (2020). How do countries compare? In *What students learn matters: Towards a 21st century curriculum*. OECD Publishing. https://doi.org/10.1787/d86d4d9a-en